

# Date Information

There are multiple dates that are relevant to how incoming files are processed through the EDH data pipeline. Here is a brief description of all of the generic date information that is tracked within the CDP.

- [Activity Date](#)
- [Date Received](#)
- [Date Created](#)
- [Date Modified](#)
- [Date Refreshed](#)

## Activity Date

- This date is recorded on each activity and represents the actual date or timestamp that the event occurred.
- This date is used internally within the pipeline to track and reconstruct the sequence in which user events took place

## Date Received

- This date represents the month or day that the information was exported by the customer and/or delivered to Influence Health.
- This is used to populate the month-year in the case of monthly or batch load
- The current design is based around monthly loads. Nightly loads would simply use the year-month-day format as the final bucket or file name format (2018-03-20/ or 2018-03-20.parquet).
- This field is not currently being tracked as an attribute on activities or within the data pipeline. It is only used for file handling on S3.

## Date Created

- This is a timestamp that is assigned when a file is run through the cleanse process and an activity parquet file is generated.
- Because of how .parquet files work: If the data is reloaded through cleanse this will update the createdAt for the affected activities
- However, if the data is reloaded from parquet into a clean database, the created date should remain the same for those activities

## Date Modified

- This is a timestamp to track when the most recent modification was applied to an Activity (or Person) within the database.
- By default this will be the same as the Date Created
- For Persons: This is caused by a new activity being identified for that person and that activity's details generating an update to our person record
- For Activities: This is typically caused by a record being sent with an existing source\_record\_id containing corrections to a previous event

## Date Refreshed

- This is a new timestamp that is being added to Redshift and Elasticsearch to track when the last refresh from the datahub (Cassandra) occurred